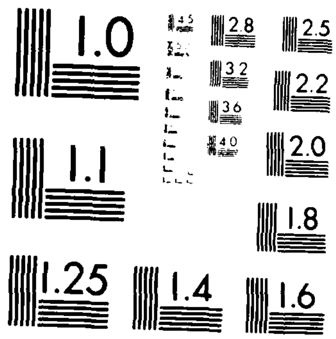
AD-A151 885 ON LIMITING EMPIRICAL DISTRIBUTION FUNCTION OF THE EIGENVALUES OF A MULTI..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Z D BAI ET AL. DEC 84 1/1

UNCLASSIFIED TR-84-42-REV AFOSR-TR-85-0262 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ON LIMITING EMPIRICAL DISTRIBUTION FUNCTION OF THE EIGENVALUES OF A MULTIVARIATE F MATRIX

Z. D. Bai[1], Y. Q. Yin[1], and
P. R. Krishnaiah[2]

Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis

University of Pittsburgh

ON LIMITING EMPIRICAL DISTRIBUTION
FUNCTION OF THE EIGENVALUES OF A
MULTIVARIATE F MATRIX

Z. D. Bai[1], Y. Q. Yin[1], and
P. R. Krishnaiah[2]

Center for Multivariate Analysis
University of Pittsburgh

(Revised version of CMA Tech. Rept. No. 84-42)

[1]Z. D. Bai and Y. Q. Yin are on a leave of absence from the China University of Science and Technology, Hefei, People's Republic of China.

[2]The work of this author was sponsored by the Air Force Office of Scientific Research (AFSC) under Contracts F49620-82-K-0001 and F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

ON LIMITING EMPIRICAL DISTRIBUTION FUNCTION OF THE EIGENVALUES OF A MULTIVARIATE F MATRIX*

Z. D. Bai, Y. Q. Yin and P. R. Krishnaiah

ABSTRACT

In this paper, the authors derived an explicit expression for the limit of the empirical distribution function (e.d.f.) of a central multivariate F matrix when the number of variables and degrees of freedom both tend to infinity in certain fashion. The authors also extended the above result to the case when the underlying distribution is not necessarily multivariate normal but the first four moments exist. The limiting distribution is useful in deriving the limiting distributions of certain test statistics which arise in multivariate analysis of variance, canonical correlation analysis and tests for the equality of two covariance matrices.

Keywords and Phrases: Canonical correlation analysis, empirical distribution function, large dimensional random matrices, limiting distribution, multivariate analysis of variance, multivariate F matrix.

*This paper is a revised version of the CMA Technical Report No. 84-42. In this version, we added Section 5 which gives some results when the underlying distribution is not multivariate normal.

A-1

1. INTRODUCTION

Various test procedures in multivariate analysis are based upon certain functions of the eigenvalues of random matrices. A considerable amount of work was done in the literature on the asymptotic distribution theory of these statistics when the sample size is very large. But, many situations arise in multivariate data analysis when the number of variables and the sample size are both very large. So, there is a great need to investigate the distributions of various functions of the eigenvalues of large dimensional random matrices. Distributions of the eigenvalues of large dimensional random matrices arise (e.g., see Mehta (1967)) in nuclear physics also.

Some work was done in the literature on the limiting empirical distribution function (e.d.f.) of large dimensional random matrices. Here, we note that the e.d.f. of a random matrix $Z: p \times p$ is defined as $N(x)/p$ where $N(x)$ denotes the number of the eigenvalues of Z which are less than or equal to x. The e.d.f. (also known as spectral distribution) of Z is useful in deriving the distributions of certain functions of the eigenvalues of Z.

Now, let $S_1 : p \times p$ be distributed as central Wishart matrix with m degrees of freedom and $E(S_1/m) = I_p$. Also, let p and m both tend to infinity such that $\lim(p/m) = y > 0$. Then, it is known (see Grenander and Silverstein (1977), Jonsson (1982) and Wachter (1978)) that the e.d.f. $F_m(x)$ of the eigenvalues of S_1/m tends to $F_y(x)$ where $F_y(x)$ is the distribution function with density function given by

$$f_y(x) = \begin{cases} \dfrac{\sqrt{(x-a)(b-x)}}{2\pi xy} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

where $a = (1-\sqrt{y})^2$, and $b = (1+\sqrt{y})^2$ and $0 < y \leq 1$; for $1 < y < \infty$, $F_y(x)$ has mass $1-(1/y)$ at zero and $f_y(x)$ on (a,b). Yin and Krishnaiah (1983b) showed that the spectral distribution of the sample covariance matrix has a limit when the underlying distribution is isotropic and $y < 1$.

Yin and Krishnaiah (1983a) showed that the spectral distribution of S_1T/m tends to a limit in probability for each x under the following conditions:

(a) T is a symmetric, positive definite matrix and $G_p(x)$ is the e.d.f. of the eigenvalues of T,

(b) S_1 and T are independent of each other,

(c) $\lim_{p,m\to\infty}(p/m) = y$ exists and finite

(d) $\int x^k dG_p(x) \to H_k$ exists in L^2 for k = 1, 2,... and $\Sigma H_{2k}^{-\frac{1}{2}k} = \infty$.

Yin and Krishnaiah (1984b) extended the above result to the case when S_1 is the sample sums of squares and cross products matrix based upon observations from an isotropic population.

Now, let S_1: $p \times p$ and S_2: $p \times p$ be distributed independently as central Wishart matrices with m and n degrees of freedom and $E(S_1/m) = E(S_2/n) = I_p$. Then, the distribution of $nS_1S_2^{-1}/m$ is known to be the central multivariate F matrix. Applying the result of Yin and Krishnaiah (1983), Yin, Bai, and Krishnaiah (1983) showed that the limit of the spectral distribution of the central multivariate F matrix exists when $p/m \to y$ exists and $(p/n) \to z < \frac{1}{2}$ as $p \to \infty$. Silverstein (1984b) showed the validity of the above result even for the case $\frac{1}{2} \leq z < 1$ by making a minor modification in the proof of Yin, Bai, and Krishnaiah (1983). Yin and Krishnaiah (1983a) gave an expression for the moments of the limit of the e.d.f. of the eigenvalues of S_1T/m. Starting from this expression, Silverstein (1984a) derived an explicit expression for the limit of the e.d.f. of the eigenvalues of the multivariate F matrix. Wachter (1980) had earlier obtained the explicit expression by using a different method. In this paper, the authors give an alternative derivation of the above method. The authors also gave explicit expressions for the moments of the above limiting distribution and these expressions are not known in the literature. Finally, the authors proved the existence of the limit of the spectral distribution of $nS_1S_2^{-1}/m$ when the underlying distribution is not multivariate normal but the first four moments exist.

2. PRELIMINARIES

In this section, we give some results which are needed in the sequel as well as a brief review of known results on limiting spectral distribution of a multivariate F matrix.

Lemma 2.1 Let $z \in (0,1)$, $a' = (1 - \sqrt{z})^2$ and $b' = (1 + \sqrt{z})^2$. If $0 \le |t| < a'$, then

$$\frac{1}{2\pi z}\int_{a'}^{b'} \frac{1}{(x - t)} \{(x - a')(b' - x)\}^{\frac{1}{2}}\, dx = \frac{1}{2z}[1 + z - t - \{(1 - z - t)^2 - 4tz\}^{\frac{1}{2}}] \tag{2.1}$$

Proof Making the transformation $u = [2x - (b' + a')]/(b' - a')$ in the left side of (2.1), we obtain

$$R(t) = \frac{1}{2\pi z}\int_{a'}^{b'} \frac{1}{(x - t)}\{(x - a')(b' - x)\}^{\frac{1}{2}}dx = \frac{(b' - a')}{4\pi z}\int_{-1}^{1} \frac{(1 - u^2)^{\frac{1}{2}}}{u + \Delta}\, du \tag{2.2}$$

where $\Delta = (b' + a' - 2t)/(b' - a')$. It is known (see Jonsson (1982)) that $R(0) = 1$. So, for any $r \in (0,1)$, we have

$$\frac{1}{\pi\sqrt{r}}\int_{-1}^{1} \frac{(1 - u^2)^{\frac{1}{2}}du}{u + ((1 + r)/2\sqrt{r})} = 1. \tag{2.3}$$

Now, let $\Delta = (1 + r)/2\sqrt{r}$. Since $\Delta > 1$, the condition $r \in (0,1)$ is satisfied. So, using (2.2) and (2.3), we obtain

$$R(t) = (b' - a')(\Delta - \sqrt{\Delta^2 - 1})/4z = [1 + z - t - \{(1 - z - t)^2 - 4tz\}^{\frac{1}{2}}]/2z. \tag{2.4}$$

Lemma 2.2 For any nonnegative integers m and w, we have

$$\sum_{?=0}^{[w/2]} (-1)^{?} \binom{m}{?}\binom{2m - 2?}{2m - w} = 2^w \binom{m}{w}. \tag{2.5}$$

Here $\binom{m}{w}$ is defined to be zero when $m < w$.

Proof If $m = 0$, the proof is trivial. We now prove the result for $m > 0$ by induction. Suppose the result is true for a fixed value of m. Then, we have

$$\sum_{\ell=0}^{[w/2]} (-1)^{\ell} \binom{m+1}{\ell} \binom{2m+2-2\ell}{2m-w+2} = \sum_{\ell=0}^{m+1} (-1)^{\ell} \left[\binom{m}{\ell} + \binom{m}{\ell-1}\right] \binom{2m+2-2\ell}{2m+2-w}$$

$$= \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{\ell} \left[\binom{2m-2\ell}{2m-w} + 2\binom{2m-2\ell}{2m+1-w}\right]$$

$$= 2^{w}\binom{m}{w} + 2 \cdot 2^{w-1}\binom{m}{w-1} = 2^{w} \binom{m+1}{w}.$$

So the result follows.

Lemma 3

$$\int \frac{(\alpha y + \beta)dy}{(\gamma + \delta y - \rho y^2)^{3/2}} = \frac{4\alpha\gamma - 2\beta\delta + (4\rho\beta + 2\alpha\delta)y}{(\delta^2 + 4\gamma\rho)(\gamma + \delta y - \rho y^2)^{1/2}} + \text{const.} \tag{2.6}$$

The above lemma can be verified directly by differentiation.

3. MOMENTS OF THE LIMITING SPECTRAL DISTRIBUTION OF A MULTIVARIATE F MATRIX

Let $F_{y,z}(x)$ denote the limit of the e.d.f. of the multivariate F matrix as defined in the preceding section. Also, let $\{E_k\}_{k=1}^{\infty}$ denote the moments of $F_{y,z}(x)$. Then, from Yin and Krishnaiah (1983), we know that

$$E_k = \sum_{w=0}^{k-1} y^w \binom{k}{w} \frac{1}{(w+1)} B(k,w) \tag{3.1}$$

where

$$B(k,w) = \sum \frac{(w+1)!}{n_1! \ldots n_{k-w}!} H_1^{n_1} \ldots H_{k-w}^{n_{k-w}} \tag{3.2}$$

and the summation in (3.2) is over all possible values of $n_1, \ldots, n_{k-w}$ subject to the restrictions $n_1 + \cdots + n_{k-w} = w + 1$ and $n_1 + n_2 + \cdots + (k-w)n_{k-w} = k$. Also, $H_i = E(x^{-i})$ for $i = 1, 2, \ldots$ where the density of x is given by

$$g_z(x) = \begin{cases} \dfrac{\sqrt{(x-a')(b'-x)}}{2\pi x z} & a' < x < b' \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where a', b', and z are as defined in the preceding section. For any $|t| < a'$,

$$\begin{aligned} \phi(t) &= tH_1 + t^2 H_2 + t^3 H_3 + \ldots . \\ &= E(t/(x-t)) \\ &= \frac{1}{2\pi z} \int_{a'}^{b'} [t\{(x-a')(b'-x)\}^{\frac{1}{2}}/(x-t)]dx \\ &= (1-z-t)[1 - \{1 - \{(4tz/(1-z-t)^2)\}^{\frac{1}{2}}\}]/2z \end{aligned} \tag{3.4}$$

by using Lemma 2.1. Since $B(k,w)$ is the coefficient of t^k in Taylor expansion of $\{\phi(t)\}^{w+1}$, we obtain

$$\begin{aligned} B(k,w) = \sum_{\ell=0}^{w+1} \binom{w+1}{\ell} (-1)^{\ell} \sum_{j=0}^{k} \binom{\ell/2}{j} (-1)^j 2^{2j-w-1} z^{j-w-1} \\ \cdot \binom{w+1-2j}{k-j} (1-z)^{w+1-j-k} (-1)^{k-j} \end{aligned} \tag{3.5}$$

If $w + 1 - 2j \geq 0$, $\binom{w+1-2j}{k-j} \neq 0$ implies $0 \leq j \leq w + 1 - k \leq 0$, i.e. $j = 0$.

Noticing $\sum_{\ell=0}^{w+1} \binom{w+1}{\ell}(-1)^{\ell} = 0$, we see that the expression of $B(k,w)$ only contains

the terms with $j > ((w+1)/2)$. If ℓ is even, then $\binom{\ell/2}{j} = 0$ implies

$j \leq \ell/2 \leq (w+1)/2$, which is contrary to $j > ((w+1)/2)$. Thus, in the expression of

$B(k,w)$, there are only the terms with $j > ((w+1)/2)$ and ℓ being odd. Applying

Lemma 2.2, we obtain

$$B(k,w) = \sum_{\ell=0}^{[w/2]} \binom{w+1}{2\ell+1}(-1)^{k+1} \sum_{j=((w+3)/2)}^{k} \binom{\ell + 1/2}{j} 2^{2j-1-w} z^{j-w-1}$$

$$\times \binom{w+1-2j}{k-j} (1-z)^{w+1-j-k}$$

$$= \sum_{\ell=0}^{[w/2]} \binom{w+1}{2\ell+1} \sum_{j=((w+3)/2)}^{k} \frac{(2\ell+1)!(2j-2-2\ell)!}{j!\ell!(j-\ell-1)!}$$

$$\times (-1)^{\ell} 2^{-w} z^{j-w-1} (1-z)^{w+1-j-k} \binom{k+j-w-2}{k-j}$$

$$= \sum_{j=w+1}^{k} \frac{(k+j-w-2)!(w+1)}{j!(j-w-1)!(k-j)!} z^{j-w-1}(1-z)^{w+1-j-k}$$

$$= \sum_{j=0}^{k-w-1} \frac{(k+j-1)!(w+1)}{(j+w+1)!j!(k-j-w-1)!} z^{j}(1-z)^{-k-j} \tag{3.6}$$

From (3.5) and (3.6) it follows that

$$E_k = \sum_{w=0}^{k-1} y^w \binom{k}{w} \sum_{j=0}^{k-w-1} \frac{(k+j-1)!}{j!(j+w+1)!(k-j-w-1)!} z^j (1-z)^{-k-j}$$

$$= \sum_{j=0}^{k-1} z^j (1-z)^{-k-j} \sum_{w=0}^{k-1-j} y^w \binom{k}{w} \frac{(k+j-1)!}{j!(j+w+1)!(k-j-w-1)!} . \tag{3.7}$$

The k-th moment of $F_{y,z}(\cdot)$ can be easily computed from (3.7)

4. DERIVATION OF THE LIMITING SPECTRAL DISTRIBUTION OF MULTIVARIATE F MATRIX

Substituting the well-known formulas

$$\binom{k}{w} = \sum_{t=0}^{w} \binom{k-j-w-1}{t}\binom{j+w+1}{w-t} \tag{4.1}$$

into (3.7), we obtain the following by changing the order of summation:

$$E_k = \sum_{j=0}^{k-1} z^j (1-z)^{-k-j} \sum_{t=0}^{[(k-1-j)/2]} \frac{(k+j-1)!\, y^t\, (1+y)^{k-j-1-2t}}{(k-j-1-2t)!\,t!\,(j+t+1)!\,j!}$$

$$= \sum_{j=0}^{k-1} z^j (1-z)^{-k-j} \sum_{t=0}^{[(k-j-1)/2]} \frac{(k+j-1)!\,y^{t-j}(1+y)^{k+j-1-2t}}{(k+j-1-2t)!\,(t-j)!\,j!\,(t+1)!}$$

$$= \sum_{t=0}^{k-1} \binom{2t}{t} \frac{1}{t+1} \sum_{j=0}^{t} z^j (1-z)^{-k-j} \binom{t}{j}\binom{k+j-1}{2t} y^{t-j}(1+y)^{k+j-1-2t} \tag{4.2}$$

Using the formulas

$$\binom{k+j-1}{2t} = \sum_{s=0}^{2t} \binom{k-1}{s}\binom{j}{2t-s} \tag{4.3}$$

in (4.2), we obtain

$$E_k = \sum_{t=0}^{k-1} \binom{2t}{t} \frac{1}{t+1} \sum_{s=t}^{2t} \binom{k-1}{s} \sum_{j=0}^{t} z^j (1-z)^{-k-j}$$

$$\times\ y^{t-j}(1+y)^{k+j-1-2t} \binom{s-t}{t-j}\binom{t}{s-t}$$

$$= \sum_{t=0}^{k-1} \binom{2t}{t} \frac{1}{t+1} \sum_{s=t}^{2t} \binom{k-1}{s}\binom{t}{s-t} \sum_{j=0}^{s-t} \binom{s-t}{j} z^{t-j}(1-z)^{-k-t+j}$$

$$\times\ y^j (1+y)^{k-t-1-j}$$

$$= \sum_{t=0}^{k-1} \binom{2t}{t} \frac{1}{t+1} \sum_{s=t}^{2t} \binom{k-1}{s}\binom{t}{s-t} z^{2t-s}(1-z)^{-k-t}$$

$$\times\ (1+y)^{k-s-1}(y+z)^{s-t}$$

$$= \sum_{t=0}^{k-1} \binom{2t}{t} \frac{1}{t+1} \sum_{s=0}^{t} \binom{k-1}{s+t} \binom{t}{s} z^{t-s} (1-z)^{-k-t}$$

$$\times (1+y)^{k-t-s-1} (y+z)^{s} \tag{4.4}$$

$$= \sum_{s=0}^{[(k-1)/2]} \sum_{t=s}^{k-1-s} \binom{k-1}{s+t} \binom{t}{s} \binom{2t}{t} \frac{1}{t+1} z^{t-s} (1-z)^{-k-t}$$

$$\times (1+y)^{k-s-1-t} (y+z)^{s}$$

$$= \sum_{s=0}^{[(k-1)/2]} \sum_{t=0}^{k-1-2s} \binom{k-1}{2s} \binom{k-1-2s}{t} \frac{(2s)!(2s+2t)!}{s!(2s+t)!(s+t+1)!}$$

$$\times z^{t} (1-z)^{-k-t-s} (1+y)^{k-2s-1-t} (y+z)^{s} \tag{4.5}$$

Define a random vector (U,V) where the marginal density of U is $\frac{2}{\pi}(1-x^2)^{-\frac{1}{2}} I_{(0,1)}(x)$ and the conditional density of V given $U = x$ is

$$\frac{|v|}{1-x^2} I_{[(1-x^2)^{1/2},((1-x^2)^{1/2}]}(v). \tag{4.6}$$

It is easy to see that

$$E\{U^{2s} V^{2t+2s}\} = \frac{(2s)!(2t+2s)!}{s!(2s+t)!(s+t+1)!} 4^{-2s-t} \tag{4.7}$$

and that

$$E\{U^{2s+1} V^{2t+2s+1}\} = 0.$$

Hence, from (4.5), (4.7) it follows that

$$E_k = E \sum_{s=0}^{k-1} \sum_{t=0}^{k-1-s} \binom{k-1}{s} \binom{k-1-s}{t} z^{t} (1-z)^{-k-t-\frac{s}{2}}$$

$$\times (1+y)^{k-s-t-1} (y+z)^{\frac{s}{2}} U^{s} V^{2t+s}$$

$$= \frac{1}{(1-z)^{k}} E\left(1 + y + \frac{4V^2 z}{1-z} + \frac{4UV\sqrt{y+z}}{\sqrt{1-z}}\right)^{k-1} \tag{4.8}$$

Now, we compute the distribution of

$$\left[1 + y + \frac{4V^2 z}{1 - z} + \frac{4UV\sqrt{y + z}}{\sqrt{1 - z}} \right] / (1 - z) . \qquad (4.9)$$

Let $w_1 = UV$ and $w_2 = V^2$. Then we can easily show that the joint distribution of (w_1, w_2) is

$$\frac{1}{\pi} w_2 (w_2 - w_1^2)^{-3/2} \; I_{[w_1^2 + w_2^2 \leq w_2]}$$

where $I_{[\alpha \leq \beta]}$ takes value 1 if $\alpha \leq \beta$ is true and zero otherwise. Applying Lemma 2.3, we can compute the density of $4[\{(y+z)^{1/2} w_1/(1-z)^{1/2}\} + \{z\, w_2/(1-z)\}]$ as follows:

Let $q_1 < q_2$ be the roots of the equation

$$\frac{(1 - z)^2}{16z^2} q^2 + \frac{1 - z}{16(y + z)} (x - q)^2 = \frac{1 - z}{4z} q \qquad (4.10)$$

in the variable q. Let $\alpha = 1$, $\beta = 0$, $\gamma = \{(1 - z)/16(y + z)\}x^2$, $\delta = \{((1-z)/4z) + ((1-z)x/8(y+z))\}$ and $\rho = (1-z)/16(y+z)$. Then

$$\Delta_1 = 4\alpha\gamma - 2\beta\delta = - \frac{1 - z}{4(y + z)} x^2,$$

$$\Delta_2 = 4\beta\rho + 2\alpha\delta = \frac{1 - z}{2z} + \frac{1 - z}{4(y + z)} x$$

$$\Delta_3 = \delta^2 + 4\gamma\rho = \frac{(1 - z)^2 [y + z + xz]}{16z^2 (y + z)} .$$

By Lemma 2.3, we get the density of $\{4w_1 (y+z)^{1/2}/(1-z)^{1/2}\} + \{4zw_2/(1-z)\}$ is as given below:

$$f_1(x) = \frac{1}{\pi} \frac{\sqrt{1 - z}}{4\sqrt{y + z}} \frac{(1 - z)^2}{16z^2} \int_{q_1}^{q_2} q \left(\frac{1 - z}{4z} q - \frac{1 - z}{16(y + z)} (x - q)^2 \right)^{-3/2} dq$$

$$= \frac{(1 - z)^{5/2}}{64\pi z^2 \sqrt{y + z}} \left. \frac{\Delta_1 + \Delta_2 q}{\Delta_3 \sqrt{\frac{(1 - z)}{4z} q - \frac{1 - z}{16(y + z)} (x - q)^2}} \right|_{q_1}^{q_2} . \qquad (4.11)$$

In proving (4.11), we use the formula $f_{x+y}(x) = \int f(x-q, q)\, dq$ where $f_{x+y}(.)$ and $f(.,.)$ are the densities of x+y and (x,y) respectively. Note that the integrand in the integral in (4.11) is zero outside the interval $[q_1, q_2]$ by the indicator factor. Since q_1 and q_2 are roots of equation (4.10), we have

$$f_1(x) = \frac{(1-z)^{5/2}}{64\pi z^2\sqrt{y+z}\,\Delta_3} \frac{(\Delta_1 + \Delta_2 q)4z}{(1-z)q} \Bigg|_{q_1}^{q_2}$$

$$= \frac{(1-z)^{5/2}\, 4z\, (-\Delta_1)\, (q_2 - q_1)}{64\pi z^2\sqrt{y+z}\,\Delta_3\,(1-z)\, q_1 q_2}. \qquad (4.12)$$

Here we note that $\frac{(1-z)}{4z} q - \frac{1-z}{16(y+z)}(x-q)^2 = (\frac{1-z}{4z} q)^2$ for $q = q_1$ or q_2.

From (4.10), we can compute

$$q_1 q_2 = \frac{z^2 x^2}{y + z - yz} \qquad (4.13)$$

$$q_2 - q_1 = \frac{2z\sqrt{(y+z)(-x^2(1-z) + 4xz + 4(y+z))}}{y + z - yz} \qquad (4.14)$$

From (4.12) - (4.14), it follows that

$$f_1(x) = \frac{(1-z)^{5/2}\, 128z^4(y+z)}{64\pi z^2\sqrt{y+z}} \frac{[x^2(1-z)/4(y+z)]\sqrt{(y+z)(-x^2(1-z) + 4xz + 4(y+z))}}{(1-z)^2[y+z+xz](1-z)\, z^2x^2}$$

$$= \frac{\sqrt{1-z}}{2\pi(y+z+xz)} \sqrt{-x^2(1-z) + 4xz + 4(y+z)}\,. \qquad (4.15)$$

From this we can easily obtain the density of $\left[1 + y + \frac{4V^2 z}{1-z} + \frac{4UV\sqrt{y+z}}{\sqrt{1-z}}\right] \Big/ (1-z)$ as

$$f_2(x) = \frac{(1-z)\sqrt{-(1-z)^2x^2 + 2(1+y+z-yz)x - (1-y)^2}}{2\pi(xz+y)}$$

$$= \frac{(1-z)^2\sqrt{(x-a)(b-x)}}{2\pi(xz+y)} \qquad (4.16)$$

where $a = \frac{(1-\sqrt{y+z-yz})^2}{(1-z)^2}$ and $b = \frac{(1+\sqrt{y+z-yz})^2}{(1-z)^2}$.

Since $f_1(x) \neq 0$ if and only if equation (4.10) has two different roots, we find by checking the steps of computation that $f_2(x) \neq 0$ if and only if $a < x < b$. Recalling (4.9), we obtain

$$E_k = \frac{1}{1-z}\int_a^b x^{k-1} f_2(x)\,dx$$

$$= \int_a^b x^k \left[\frac{f_2(x)}{x(1-z)}\right] dx. \qquad (4.17)$$

Now, let

$$f(x) = \frac{f_2(x)}{x(1-z)} = \begin{cases} \dfrac{(1-z)\sqrt{(x-a)(b-x)}}{2\pi x(y+xz)} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases} \quad (4.18)$$

In [8], it is shown that the distribution $F_{y\cdot z}$ is determined by all its moments. From (4.11),(4.12) it follows that

$$F_{y\cdot z}(x) = \gamma I_{\{(0,\infty)\}}(x) + \int_{-\infty}^{x} f(x)dx, \quad (4.19)$$

where $\gamma = 1 - \int_{-\infty}^{\infty} f(x)dx$.

Finally, we only need to compute the integral $\int_a^b f(x)dx$. If we set $u = [2x - (b+a)]/(b-a)$, we get

$$I = \int_a^b f(x)dx = \frac{(1-z)}{2\pi y}\left[\int_a^b \frac{\sqrt{(x-a)(b-x)}}{x}dx - \int_a^b \frac{z\sqrt{(x-a)(b-x)}}{(z(x+y))}dx\right]$$

$$= \frac{(1-z)(b-a)}{4\pi y}\left[\int_{-1}^{1} \frac{\sqrt{1-u^2}}{u+\lambda_1}du - \int_a^b \frac{\sqrt{1-u^2}}{u+\lambda_2}du\right],$$

where $\lambda_1 = \frac{b+a}{b-a}$ and $\lambda_2 = \frac{b+a+(2y)/z}{b-a}$. Using (2.3) we get

$$I = \frac{(1-z)(b-a)}{4y}\left[(\lambda_1 - \sqrt{\lambda_1^2 - 1}) - (\lambda_2 - \sqrt{\lambda_2^2 - 1})\right]$$

$$= \frac{(1-z)}{2y}\quad [(z^2ab + zy(a+b) + b^2)^{1/2}/z - \sqrt{ab} - (y/z)]$$

$$= \frac{(1-z)}{2y}\left[\frac{y+z}{z(1-z)} - \frac{|1-y|}{1-z} - \frac{y}{z}\right]$$

$$= \frac{1}{2y}[1 + y - |1-y|]$$

$$= \begin{cases} 1 & \text{if } 0 < y \le 1. \\ \dfrac{1}{y} & \text{if } y > 1. \end{cases}$$

Hence

$$\gamma = \begin{cases} 0 & \text{if } 0 < y \le 1. \\ 1 - \dfrac{1}{y} & \text{if } y \ge 1. \end{cases}$$

Substituting this into (4.19), we get the expression of $F_{y\cdot z}(x)$.

5. LIMITING DISTRIBUTION WHEN THE UNDERLYING DISTRIBUTION IS NOT MULTIVARIATE NORMAL

In this section, we prove the existence of the limiting spectral distribution of the random matrix F when the underlying distribution is not multivariate normal but the first four moments exist. We need the following difinition and lemmas in the sequel.

Definition 5.1. A random vector $\underset{\sim}{x}' = (x_1,\ldots,x_p)$ is said to be M - PD (projection distribution) bounded if there is a positive constant M such that for any constant unit vector $\underset{\sim}{a}' = (a_1,\ldots,a_p)$ and any $x > 0$

$$P[|\underset{\sim}{a}'\underset{\sim}{x}| \leq x] \leq Mx \tag{5.1}$$

Definition 5.2. A sequence of random vectors $\underset{\sim}{x}_n$ of order $p(n) \times 1$ is said to be M - PD bounded if there is a common constant M such that each vector $\underset{\sim}{x}_n$ is M - PD bounded.

Example 5.1. If $\underset{\sim}{x}$ is distributed as a multivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix I, then $\underset{\sim}{x}$ is $\sqrt{\frac{2}{\pi}}$ - PD bounded.

Example 5.2. Let $\underset{\sim}{x}$ be uniformly distributed on the p-dimensional sphere with center at zero and radius $\sqrt{p}$, $p \geq 3$. Then $\underset{\sim}{x}$ is 2 - PD bounded.

Lemma 5.1. Let $\underset{\sim}{x}$ be an M - PD bounded random vector and let y be a random variable with $E(1/|y|) < \infty$. In addition, we assume that y and $\underset{\sim}{x}$ are distributed independent of each other. Then $y\underset{\sim}{x}$ is $ME(1/|y|)$ - PD bounded.

The proof of the above lemma follows immediately.

Lemma 5.2. If A: $p \times p$ is any nonnegative definite matrix and $\bar{\lambda}$ is the largest eigenvalue of A, then

$$|\underset{\sim}{x}'A\underset{\sim}{x} - \underset{\sim}{y}'A\underset{\sim}{y}| \leq \bar{\lambda}\ ||\underset{\sim}{x}-\underset{\sim}{y}||^2 + 2\bar{\lambda}^{1/2}\ ||\underset{\sim}{x}-\underset{\sim}{y}||\ \ ||\underset{\sim}{y}'A\underset{\sim}{y}||^{1/2} \tag{5.2}$$

where $||\underset{\sim}{x}||$ denotes the Eucledian norm of $\underset{\sim}{x}$.

The above lemma can be proved by applying Schwarz's inequality.

Lemma 5.3. If $0 < r < 1/2$, then the unit ball in R^p can be covered by balls of radius r in such a way that the number of these smaller balls does not exceed $C \exp\{(p/2) \log(8\pi e/r^2)\}$ where C is a constant.

The proof of the above lemma is given in Yin, Bai and Krishnaiah (1983).

Lemma 5.4. Let $\{\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_p,\ldots\}$ be a sequence of random vectors which are M - PD bounded with a common M and $\underset{\sim}{x}_p$ is of order $p \times 1$. Also, let $\underset{\sim}{x}_{p1},\ldots,\underset{\sim}{x}_{pn}$ be a sample of $n = n(p)$ independent observations on $\underset{\sim}{x}_p$. Then, for any $\underset{\sim}{a}$: $p \times 1$ and $\varepsilon > 0$

$$P\left[\sum_{i=1}^{n} (\underset{\sim}{a}'\underset{\sim}{x}_{pi})^2 \leq n\varepsilon\right] \leq C \exp\{\frac{n}{2} \log (\pi e M^2 \varepsilon/2)\} \tag{5.3}$$

Proof. Let $Y_i = \underset{\sim}{a}'\underset{\sim}{x}_{pi}$ for $i = 1,2,\ldots,n$. Then $Y_1,\ldots,Y_n$ are i.i.d. random variables and

$$V(x) = P[|Y_1| \leq x] \leq Mx \tag{5.4}$$

But, we know that

$$P[Y_1^2 + \ldots + Y_n^2 \leq n\varepsilon] \leq P\left[\sum_{k=1}^{[n/2]} \{Y_{2k-1}^2 + Y_{2k}^2\} \leq n\varepsilon\right] \tag{5.5}$$

where [a] is an integral part of a. We have

$$\begin{aligned} P[Y_1^2 + Y_2^2 \leq x] &= \int_0^x V(\sqrt{x-y})\,dV(\sqrt{y}) \\ &\leq M \int_0^x \sqrt{x-y}\; dV(\sqrt{y}) \\ &\leq \frac{M^2}{2} \int_0^x \sqrt{y/(x-y)}\; dy. \end{aligned} \tag{5.6}$$

But

$$\int_0^x \{y/(x-y)\}^{1/2}\, dy = \frac{\pi x}{2} \tag{5.7}$$

So

$$P[Y_1^2 + Y_2^2 \leq x] \leq \frac{\pi M^2}{4} x. \tag{5.8}$$

Now, let $Z_k = Y_{2k-1}^2 + Y_{2k}^2$. Then $Z_1, \ldots, Z_a$ are i.i.d. nonnegative random variables $a = [n/2]$ and

$$P[Z_1 \leq x] \leq M^* x \tag{5.9}$$

where $M^* = \pi M^2/4$. Now let

$$F_a(x) = P[Z_1 + \ldots + Z_a \leq x]. \tag{5.10}$$

We shall now prove, by induction, that

$$F_a(x) \leq M^a x^a / a! \ . \tag{5.11}$$

The above result is true for $a = 1$. Here and in (5.12) below, we use M instead of M^* for simplicity. We will assume the result to be true for $a = b$.

Now
$$\begin{aligned} F_{b+1}(x) &= \int_0^x F_1(x-y)\,dF_b(y) \\ &\leq M \int_0^x (x-y)\,dF_b(y) \\ &= M \int_0^x F_b(y)\,dy \\ &\leq \frac{M^{b+1}}{b!} \int_0^x y^b \, dy = \frac{M^{b+1} x^{b+1}}{(b+1)!} \end{aligned} \tag{5.12}$$

So, by induction, the inequality (5.11) holds good. Now, using inequalities (5.5), (5.8) and (5.11), we obtain

$$P[Y_1^2 + \ldots + Y_n^2 \leq n\varepsilon] \leq (\pi \ M^2/4)^a \ (n\varepsilon)^a / a! \ . \tag{5.13}$$

Now, by applying Stirling's formula, we obtain the desired result.

We now prove the following result

<u>Theorem 5.1</u>. Let $\underset{\sim}{X}_p$, $p = 1,2,\ldots$ be M - PD bounded and $(p/n) \to y \ \varepsilon \ (0,1)$. Then, for any $R > 0$, there exist positive constants $C < \infty$, $D < \infty$, $\alpha > 0$, $\varepsilon_0 > 0$ depending only on y, M and R, such that

$$P[\varepsilon \leq \underline{\lambda}_p \leq \overline{\lambda}_p \leq R] \leq CD^p \varepsilon^{\alpha p}, \quad 0 < \varepsilon \leq \varepsilon_0$$

where $\underline{\lambda}_p$ and $\overline{\lambda}_p$ are the smallest and largest eigenvalues of W_p

respectively where

$$W_p = \frac{1}{n} \sum_{j=1}^{n} X_{pj} X'_{pj} \tag{5.14}$$

and $X_{p1}, \ldots, X_{pn}$ are independent observations on X_p.

Proof. Let S_p denote the unit sphere in R_p. Then,

$$\underline{\lambda}_p = \min_{y \in S_p} y' W_p y. \tag{5.15}$$

Let $r < 1/2$ be a positive constant to be chosen later. According to Lemma 5.3, let $B_p(x_1, r), \ldots, B_p(x_q, r)$ be those balls with radius r and centers $x_1, \ldots, x_q$ which cover the unit sphere S_p. In addition, q satisfies

$$q \leq C \exp\{\frac{p}{2} \log \frac{8\pi e}{r^2}\} . \tag{5.16}$$

Then we have

$$P(\underline{\lambda}_p \leq \varepsilon, \ \overline{\lambda}_p \leq R) \leq \sum_{k=1}^{q} P(\min_{y \in B(x_k, r) \cap S_p} y' W_p y \leq \varepsilon, \overline{\lambda}_p \leq R).$$

If $y \in B(x_k, r) \cap S_p$ and $y' W_p y \leq \varepsilon$, $\overline{\lambda}_p \leq R$, then by Lemma 5.2, we have

$$\begin{aligned} x_k' W_p x_k &\leq y' W_p y + |x_k' W_p x_k - y' W_p y| \\ &\leq \varepsilon + R r^2 + 2R^{1/2} r \varepsilon^{1/2} = (\varepsilon^{1/2} + R^{1/2} r)^2. \end{aligned}$$

Therefore,

$$P(\underline{\lambda}_p \leq \varepsilon, \overline{\lambda}_p \leq R) \leq \sum_{k=1}^{q} P(x'_k W_p x_k \leq (\varepsilon^{1/2} + R^{1/2} r)^2)$$

But by Lemma 5.4, we have

$$\begin{aligned} &P(x_k' W_p x_k \leq (\varepsilon^{1/2} + R^{1/2} r)^2) \\ &\quad \leq C \exp\{\frac{n}{2} \log(\frac{\pi e}{2} M^2 (\varepsilon^{1/2} + R^{1/2} r)^2/(1-r)^2)\} \end{aligned} \tag{5.17}$$

If we take $r = R^{-1/2} \varepsilon^{1/2}$, we have from (5.16) and (5.17),

$$P(\lambda_p \le \varepsilon, \bar{\lambda}_p \le R)$$

$$\le C \exp\{\frac{p}{2} \log \frac{8\pi eR}{\varepsilon} + \frac{n}{2} \log(8\pi e M^2 \varepsilon)\}$$

$$\le C \varepsilon^{\alpha p} D^p$$

for p large enough where $D = (8\pi eR)^{1/2}(8\pi M^2)^{1/y}$, $0 < \alpha < 1/2(1/y - 1)$. But (5.17) holds provided that $r = R^{-1/2}\varepsilon^{1/2} < 1/2$, i.e. $\varepsilon < \varepsilon_0 \triangleq \frac{1}{4} R$. So, the theorem is proved.

Theorem 5.2. Suppose that the assumptions of Theorem 5.1 hold. Also, we assume that $\underset{\sim}{X}'_{pi} = (X_{i1},\ldots,X_{ip})$, and X_{ij}, $i,j = 1,2,\ldots$, are i.i.d. random variables with common mean zero and variance σ^2 and finite fourth moment. Then there exist a positive number ε such that

$$\varliminf_{p\to\infty} \lambda_p \ge \varepsilon, \quad \text{a.s.}$$

Proof. Take $R > (1+\sqrt{y})^2\sigma$ and set $A_p = \bigcup_{m=p}^{\infty} (\bar{\lambda}_m \ge R)$. According to Yin, Bai and Krishnaiah (1984), $P(A_p) \to 0$ as $p \to \infty$. Thus for any $\varepsilon < \varepsilon_0$, we have

$$P(\bigcup_{m\ge p}^{\infty} (\lambda_m < \varepsilon)) \le \sum_{m=p}^{\infty} P(\lambda_m \le \varepsilon, \bar{\lambda}_m \le R) + P(A_p)$$

$$\le C \sum_{m=p}^{\infty} (\varepsilon^{\alpha} D)^p + P(A_p) \to 0, \; p \to \infty.$$

We will now prove that the limiting spectral distribution of F matrix exist under conditions weaker than assumed in Sections 3 and 4.

Theorem 5.3. Let $W_p = \frac{1}{n} X_p X_p'$, $\tilde{W}_p = \frac{1}{m} Y_p Y_p'$ where $X_p = (X_{ij})$: $p \times n$, $Y_p = (Y_{ij})$: $p \times m$ where $n = n(p)$, $m = m(p)$ such that $(p/n) \to y \in (0,\infty)$, $(p/m) \to y' \in (0,1)$. Also, we assume that X_{ij}, Y_{it} ($i = 1,2,\ldots,p$; $j = 1,2,\ldots,n$; $t = 1,2,\ldots,m$) are i.i.d. random variables with common mean zero, variance one and finite fourth moment. Also, let $F_p = W_p \tilde{W}_p^{-1}$ where $\tilde{W}_p$ is nonsingular. Also, for each p, we assume that

$(Y_{11},\ldots,Y_{p1})$ is M - PD bounded. Then, as $p \to \infty$, F_p has a limiting spectral distribution with probability one.

Proof. Applying Theorem 5.2, we have

$$0 < \varepsilon \le \underline{\lambda}_p \le \overline{\lambda}_p \le R \triangleq (1+\sqrt{y})^2 + 1$$

for sufficiently large P, with probability one. Thus, for any $k \ge 1$,

$$\frac{1}{p}\,\mathrm{tr}(\tilde{W}_p)^k = \int_0^\infty x^{-k}dQ_p(x) = \int_\varepsilon^R x^{-k}dQ_n(x)$$

where Q_p is the e.d.f. of $\tilde{W}_p$.

According to Yin (1984), the random matrix W_pT has a limiting spectral distribution if W_p satisfies the condition of the theorem and T satisfies the conditions (a), (b), (c) stated on page 2 and the following condition:

$$(d') \int x^k dQ_p(x) \to H_k \quad \text{a.s. for } k = 1,2,\ldots \text{ and } \Sigma\, H_{2k}^{-\frac{1}{2}k} = \infty.$$

Taking T = I, the identity, we know that $Q_p \xrightarrow{W} Q$, a.s. where Q is the distribution whose density is given by (3.3). Thus

$$\frac{1}{p}\,\mathrm{tr}(\tilde{W}_p)^k \longrightarrow H_k = \int_{a'}^{b'} x^{-k} g_z(x)dx, \quad \text{a.s.}$$

Again, using Yin (1984), with $T = (\tilde{W}_p)^{-1}$, the theorem is proved. In view of Yin (1984) and Yin and Krishnaiah (1983a) and the derivation in §4, we know that the limiting spectral distribution is the same as that given in §4.

REFERENCES

[1] GRENANDER, U. and SILVERSTEIN, J.W. (1977). "Spectral analysis of networks with random topologies". SIAM J. Appl. Math., 32, 499-519.

[2] JONSSON, D. (1982). "Some Limit theorems for the eigenvalues of a sample covariance matrix". J. Multivariate Anal., 12, 1-38.

[3] MEHTA, M.L. (1967). Random Matrices and the Statistical Theory of Energy Levels. Academic Press, Inc.

[4] SILVERSTEIN, J.W. (1984a). "The limiting eigenvalue distribution of a multivariate F matrix". To appear in SIAM J. Math. Anal.

[5] SILVERSTEIN, J.W. (1984b). "Comments on a result of Yin, Bai and Krishnaiah". J. Multivariate Anal., 15.

[6] WACHTER, K.W. (1978). "The strong limits of random spectra for sample matrices of independent elements". Ann. Prob., 6, 1-18.

[7] WACHTER, K.W. (1980). "The limiting empirical measure of multiple discriminant ratios". Ann Statist., 8, 937-957.

[8] YIN, Y.Q., BAI, Z.D. and KRISHNAIAH, P.R. (1983). "Limiting behavior of the eigenvlaues of a multivariate F matrix". J. Multivariate Anal., 13, 508-516.

[9] YIN, Y.Q. and KRISHNAIAH, P.R. (1983a). "A limit theorem for the eigenvalues of product of two random matrices". J. Multivariate Anal., 13, 489-507.

[10] YIN, Y.Q. and KRISHNAIAH, P.R. (1983b). "Limit theorems for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic". To appear in Teoriya Veroyatnostei iee Primeneniya.

[11] YIN, Y.Q. and KRISHNAIAH, P.R. (1983c). "Limit theorems for the eigenvalues of product of large dimensional random matrices when the underlying distribution is isotropic". To appear in Teoriya Veroyatnostei i e e Primeneniya.

[12] YIN, Y.Q. (1984). Unpublished manuscript.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Revised Version of CMA Tech. Report No. 84-42	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 85-0262
6a. NAME OF PERFORMING ORGANIZATION University of Pittsburgh / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Center for Multivariate Analysis Pittsburgh PA 15260	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0008
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT NO.

11. TITLE (Include Security Classification)
ON LIMITING EMPIRICAL DISTRIBUTION FUNCTION OF THE EIGENVALUES OF A MULTIVARIATE F MATRIX

12. PERSONAL AUTHOR(S)
Z.D. Bai, Y.Q. Yin, and P.R. Krishnaiah

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	DEC 84	20

16. SUPPLEMENTARY NOTATION
Z.D. Bai and Y.Q. Yin are on a leave of absence from the China University of Science and Technology, Heifei, People's Republic of China.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Canonical correlation analysis; empirical distribution function; large dimensional random matrices; limiting distribution; multivariate analysis of variance; (CONT.)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this paper, the authors derived an explicit expression for the limit of the empirical distribution function (e.d.f.) of a central multivariate F matrix when the number of variables and degrees of freedom both tend to infinity in certain fashion. The authors also extended the above result to the case when the underlying distribuiton is not necessarily multivariate normal but the first four moments exist. The limiting distribution is useful in deriving the limiting distributions of certain test statistics which arise in multivariate analysis of variance, canonical correlation analysis and tests for the equality of two covariance matrices.

ITEM #18, SUBJECT TERMS, CONTINUED: multivariate F matrix.

NOTE (REF ITEM #4): In this version of CMA Technical Report No. 84-42, the authors added Sec. 5 which gives some results when the underlying distribution is not multivariate normal.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767- 5027 / 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

END

FILMED

5-85

DTIC